## Artificial

## Intelligence and Machine Learning

Project Report

Semester-IV (Batch-2022)

**Case Study** – Amazon Purchases Dataset

URL: <https://drive.google.com/file/d/1sXbd1N4KF6c89mHCxD7VZ9Ulj_giaZqY/view?usp=sharing>

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Bhardwaj Sukoon, 2210990870 (g-13)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Description:**

1. Display Top 10 rows.
2. Display Last 10 row.
3. Check Datatype of Each Column.
4. Check null values in the DataSet.
5. Number of Rows and Columns in the Dataset.
6. Highest and lowest purchase price in dataset.
7. Average purchase price.
8. How many people have French 'fr' as their language.
9. Job title contains engineer.
10. Find email of the person with the following IP address: 132.207.160.22
11. Find Email by Credit Card Number (4105595335494659).
12. People Purchasing During AM and PM.
13. Credit Cards Expiring in 2020.
14. Top 5 Most Popular Email Providers.

**Libraries:**

1. Pandas

**Methods:**

1. Display Top 10 Rows:-   
   Use the head() function in Python (Pandas) or a similar function in other data analysis tools.
2. Display Last 10 Rows:  
   Use the tail() function in Python (Pandas) or a similar function in other data analysis tools.
3. Check Datatype of Each Column:  
   Utilize the dtypes attribute in Python (Pandas) or similar functions in other tools.
4. Check Null Values in the Dataset:-   
   Use isnull() or info() function in Python (Pandas) or similar functions in other tools.
5. Number of Rows and Columns in the Dataset: Use the shape attribute in Python (Pandas) or similar functions in other tools.
6. Highest and Lowest Purchase Price in Dataset:  
   Use max() and min() functions on the Purchase Price column in Python (Pandas) or similar functions in other tools.
7. Average Purchase Price:  
   Use mean() function on the Purchase Price column in Python (Pandas) or similar functions in other tools.
8. People with French as Their Language:  
   Filter the dataset based on the 'Language' column where the value is 'fr'.
9. Job Title Contains 'Engineer':  
   Filter the dataset based on the 'Job Title' column for entries containing the term 'Engineer'.
10. Find Email by IP Address (132.207.160.22):  
    Search the dataset for the row where the IP address matches and retrieve the corresponding email.
11. Mastercard Holders with Purchases Above 50:  
    Filter the dataset for entries with 'Mastercard' as the Credit Card Provider using `contains` and 'Purchase Price' above 50.
12. Find Email by Credit Card Number (4105595335494659):  
    Search the dataset for the row where the Credit Card number matches and retrieve the corresponding email.
13. People Purchasing During AM and PM:  
    Use the time-related information in the dataset to categorize purchases as AM or PM and count accordingly.
14. Credit Cards Expiring in 2020:  
    Extract the year from the 'Credit Card Expiry' column, and check if contains “/20”
15. Top 5 Most Popular Email Providers:  
    Extract the email domain from the 'Email' column, split by “@”, use `1` index and use value\_counts() or similar functions to find the top 5.